Responsible AI licenses (RAIL)

Critical Science Reading Group
13 June 2025

Motivation

- AI tools have dual-use risk, e.g., a generative AI model trained for drug discovery can be misused to generate biochemical weapons[1]

- Open sharing is vital for scientific progress, and allows decentralization, accessibility, replicability, etc. Some publishing venues even require it.

- Traditional open-source licenses (e.g., MIT, Apache 2.0, CC) allow unrestricted use, including harmful applications, and are not adapted for ML Models and Data.

- Developers have no control over how their ML/AI software is used once it is released publicly, especially if open-sourced

[1] Urbina, F., Lentzos, F., Invernizzi, C., & Ekins, S. (2022). Dual Use of Artificial Intelligence-powered Drug Discovery. *Nature machine intelligence, 4*(3), 189–191. https://doi.org/10.1038/s42256-022-00465-9

Why licenses?

- Ethical AI guidelines (e.g., <u>Montréal Declaration on Responsible AI</u>, <u>OECD AI Principles</u>) are useful and help shape policy, but not directly enforceable

- Government regulation tends to lag behind development of new technologies (especially true for AI), leaving a gap in the short term.

- Unique challenges of AI systems: lack reliable control and predictability, their behavior depends heavily on training data and model design.

- Need to specific capabilities and limitations of an AI model, which are difficult to encode in a legislation with general rules, and even harder to apply these rules.

How can licenses help?

- **Legally enforceable** via existing contractual and IP law

- **Serve as deterrent:** large well-lawyered corporations & organizations comply to avoid the risk of expensive litigation and/or reputation harm.

- **Norm-building:** RAIL licenses promote a culture of responsible AI sharing that respects model limitations and the values of its creators.

RAILs are a readily available tool that developers can use *today* to take a step towards responsible AI, as part of a broader system of AI governance tools and not as the only solution.

What is RAIL?

- **Goal:** Empower developers to restrict the use of their AI technology in order to prevent irresponsible and harmful applications

- **RAIL:** license for AI artifacts which

  1. include a list of behavioral-use restrictions

  2. requires downstream derivatives (as defined in the license; can include fine-tuned and distilled models) to include at least the same use restrictions.

- **RAIL initiative:** a volunteer driven initiative established in 2019. Their steering committee includes AI researchers and intellectual property lawyers.

RAIL variants

- **RAIL:** license for AI artifacts which

  1. include a list of behavioral-use restrictions

  2. requires downstream derivatives (as defined in the license; can include fine-tuned and distilled models) to include at least the same use restrictions.

- RAIL can be used to license data (D), Apps (A), models (M), and source code (S).

- License can be proprietary, restricted to research use, or open (OpenRAIL)

What is OpenRAIL?

- **OpenRAIL:** subclass of RAIL licenses that permit free-of-charge open access, re-use, distribution, and commercialization

- OpenRAIL are *not* open source licenses based on Open Source Initiative definition (criteria 6) nor a Free Software (Freedom 0)
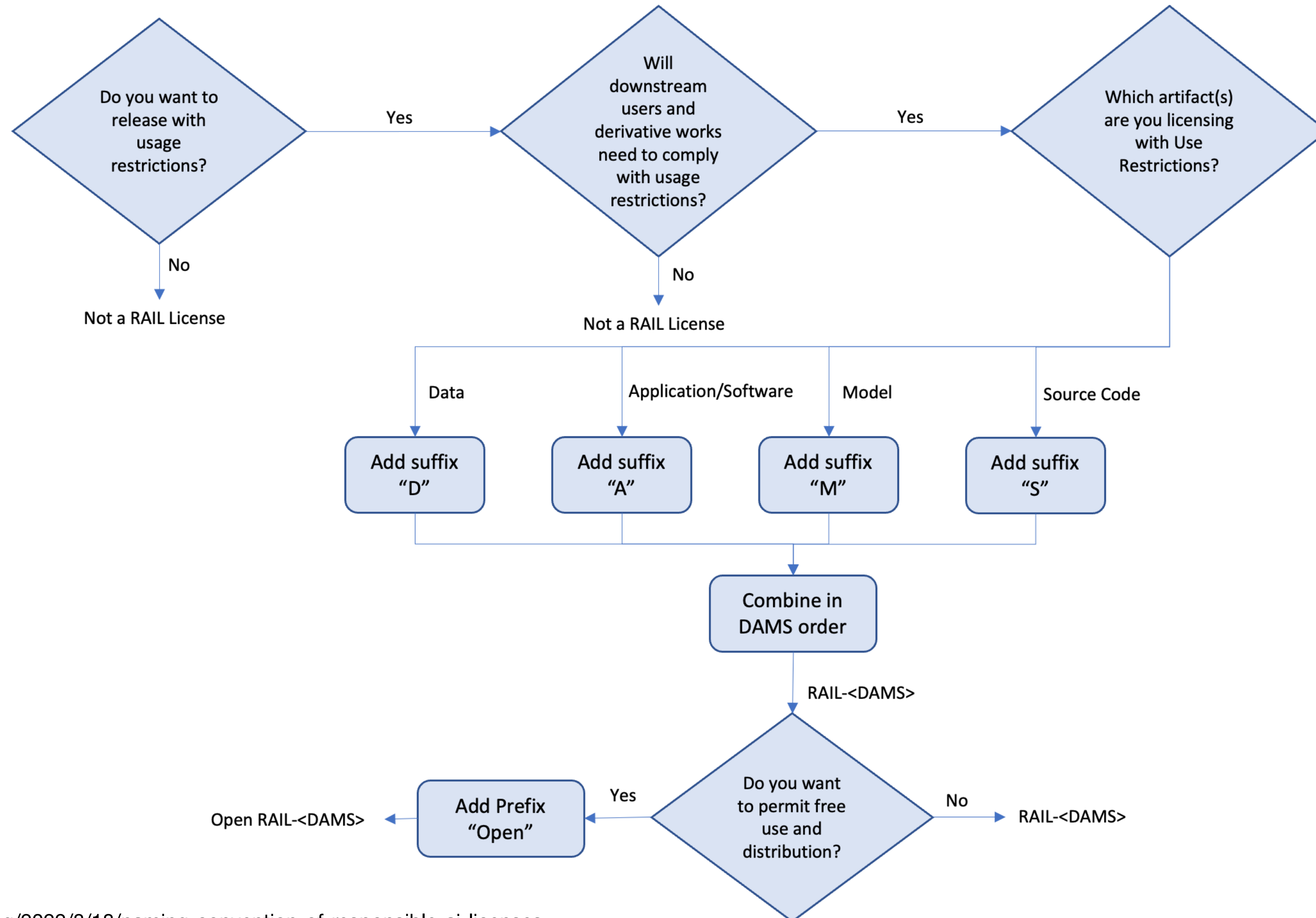
- OpenRAIL are *not* creative commons licenses

OpenRAIL licenses could be for AI artifact what open software licenses are to code and Creative Commons to general content

Comparison to Open Source

License	Licensor permits modification and redistribution	Licensor requires source code be disclosed when re-used	Licensee must include copyright notice	Licensor includes Use Restrictions
GNU Affero General Public License v3.0	Yes	Yes	Yes	No (OSI)
Apache 2.0	Yes	No	Yes	No (OSI)
Creative Commons Attribution Share Alike 4.0	Yes	No	Yes	No (CC)
Creative Commons Zero 1.0 Universal	Yes	No	No	No (CC)
MIT License	Yes	No	Yes	No (OSI)
RAIL Licenses	May or May Not	May or May Not	Yes	Yes
OpenRAIL-D	Yes	N/A	N/A	Yes
OpenRAIL-A	Yes	No	N/A	Yes
OpenRAIL-M	Yes	No	Yes	Yes
OpenRAIL-S	Yes	No	Yes	Yes

Source: https://www.licenses.ai/blog/2022/8/18/naming-convention-of-responsible-ai-licenses

RAIL naming convention

Source: https://www.licenses.ai/blog/2022/8/18/naming-convention-of-responsible-ai-licenses

Common Use restrictions

- Discrimination

- Weaponry and warfare

- Surveillance

- Disinformation

- Privacy violation

- Health care, e.g. to predict the likelihood of filing an insurance claim, diagnose a medical condition without human oversight.

Behavioral Restrictions	AIPubs RAIL	BigSci. OpenRAIL	CodeML OpenRAIL	LLaMA 2	FALCON	ImpACT L/M/H	GRID
Discrim.							
(1) Discriminate against people based on legally protected characteristics	✓	✓	✓	✓			✓
(2) Administrate justice, law enforcement, immigration, or asylum processes, such as predicting that a natural person will commit a crime.		✓	✓			✓	✓
(3) Defame, disparage or otherwise harass others.	✓	✓	✓	✓	✓		✓
(4) Exploiting, harming or attempting to exploit or harm minors in any way	✓	✓	✓	✓	✓		✓
Disinformation							
(5) Create, present or disseminate verifiably false or misleading information for economic gain, harm people, or to intentionally deceive the public.	✓	✓	✓	✓			✓
(6) Synthesize or modify a natural person's appearance, voice, or other individual characteristics, unless prior informed consent of said person is obtained.	✓	✓	✓	✓			✓
(7) Generate/disseminate information and place the information in any public context without expressly disclaiming that the content is machine generated.	✓	✓	✓	✓		✓	✓
Legal							
(8) Fully automate decision-making that creates, modifies or terminates a binding, enforceable obligation between entities.	✓	✓	✓			✓	✓
(9) Violate any national, federal, state, local or international law or regulations.	✓	✓	✓	✓	✓		✓
Privacy							
(10) Utilize personal information to infer additional personal information, including legally protected characteristics, vulnerabilities or categories; unless informed consent from the data subject is received.				✓			
(11) Generate/disseminate personal identifiable information that can be used to harm an individual or to invade the personal privacy of an individual.	✓	✓	✓	✓			✓
(12) Engage in, promote, incite, or facilitate the harassment, abuse, threatening, or bullying of individuals or a group.	✓	✓	✓	✓	✓		✓
Health							
(13) Provide medical advice or clinical decisions without accreditation of the system; unless the use is (i) in an internal research context with independent and accountable oversight and/or (ii) with medical professional oversight.		✓	✓		✓		
(14) In connection with any activities that present a risk of death or bodily harm to individuals, including self-harm or harm to others, or in connection with regulated or controlled substances.				✓			
(15) In connection with activities that present a risk of death or bodily harm to individuals, including inciting or promoting violence, abuse, or any infliction of bodily harm.				✓			
(16) Exploitaion of vulnerabilities of a group/person in a manner that is likely to cause physical or psychological harm	✓	✓	✓				
Military							
(17) Use for active deployment of weaponry.							✓
(18) Use for purposes of building or optimizing military weapons or in the service of nuclear proliferation or nuclear weapons technology.				✓	✓		
(19) Use for purposes of military surveillance, including any research related to military surveillance.				✓	✓		
(20) 'Real time' remote biometric processing or identification systems in publicly accessible spaces for the purpose of law enforcement.						✓	✓
Other							
(21) Generate/disseminate malware/ransomware or other content for the purpose of harming electronic systems.			✓				
(22) Fail to appropriately disclose to end users any known dangers of your AI system			✓				

Table 1. **Summary of Behavioral-Use Clauses.** Clauses included in popular responsible AI licenses.

Source: McDuff, D., Korjakow, T., Cambo, S., Benjamin, J. J., Lee, J., Jernite, Y., ... & Contractor, D. (2024). On the standardization of behavioral use clauses and their adoption for responsible licensing of ai. arXiv preprint arXiv:2402.05979.
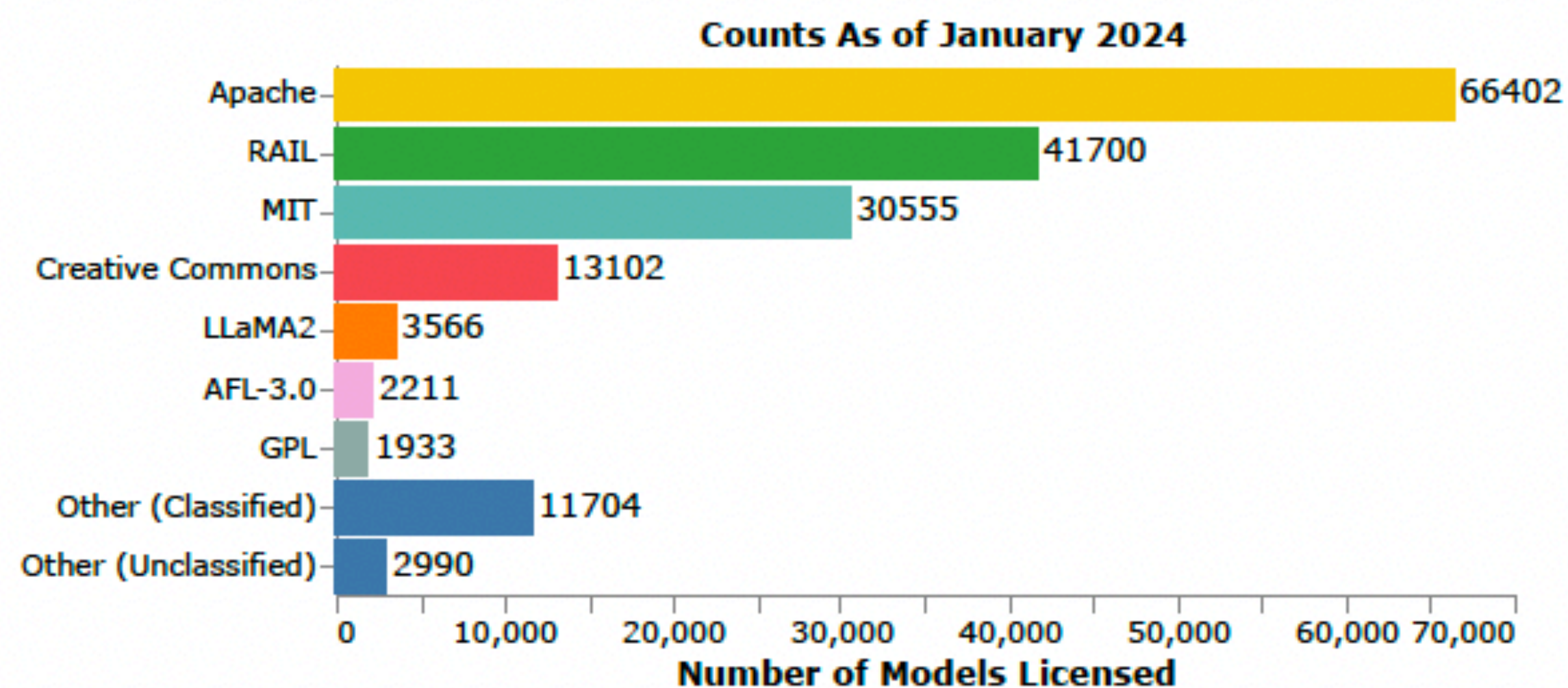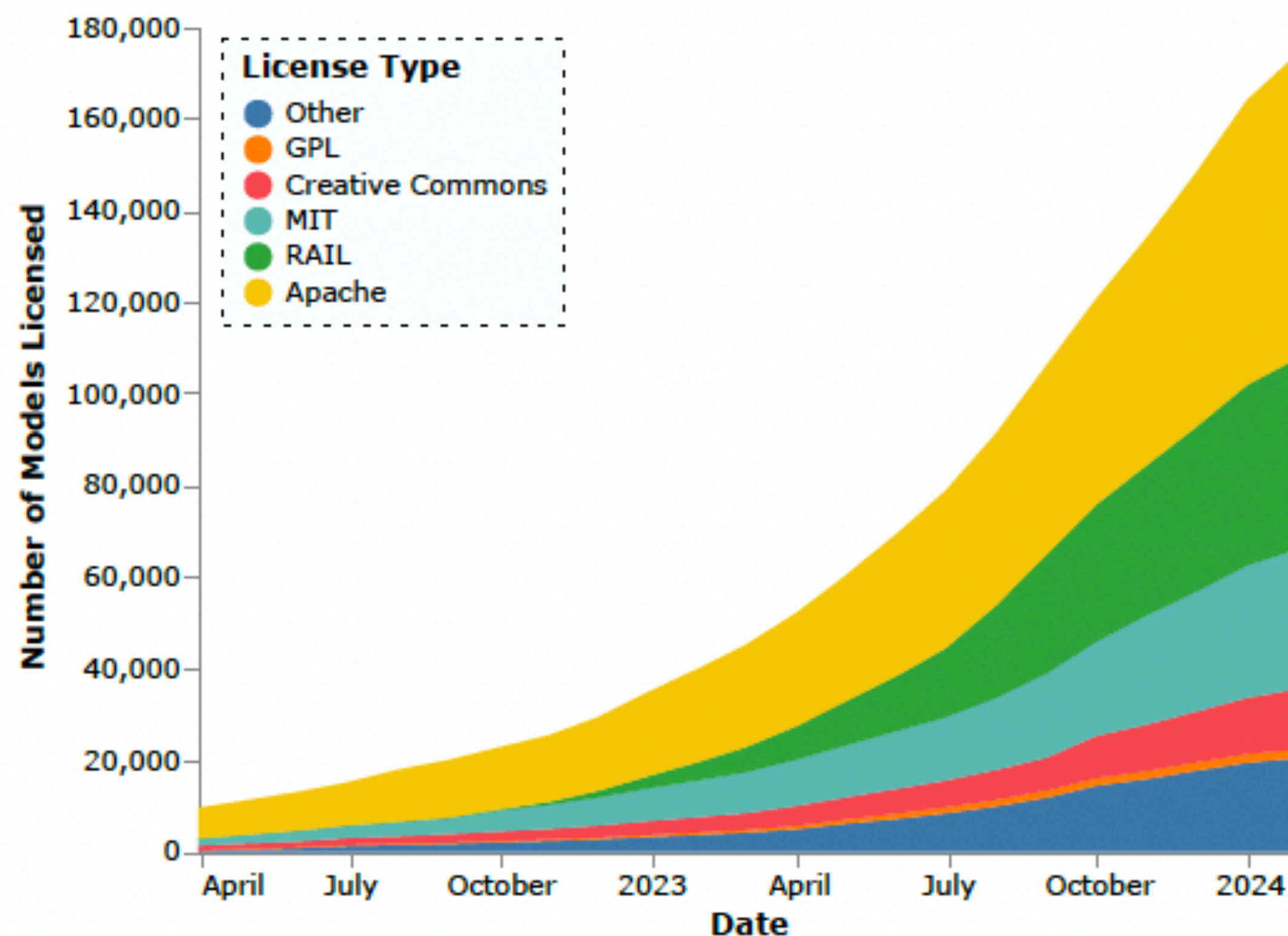
Who is using RAIL?

Examples of famous AI models using RAIL:

- BLOOM model by BigScience uses *BigScience BLOOM RAIL 1.0*, which is the first OpenRAIL-M license. Later updated to *BigScience OpenRAIL-M* to apply to any model.

- Stable Diffusion by Stability.ai uses *CreativeML OpenRAIL-M*

- BigCode by Hugging Face & ServiceNow uses *CodeML OpenRAIL-M 0.1*

- Llama models by Meta all use RAIL licenses, but not OpenRAIL, e.g., limits use by entities with >700 million monthly users unless separately licensed

Who is using RAIL?

Total # of models on Hugging Face: 174,163

Most common licenses: Apache (38.1%), RAIL10 (24.0%), MIT (17.5%), CC (7.5%)

Source: McDuff, D., Korjakow, T., Cambo, S., Benjamin, J. J., Lee, J., Jernite, Y., ... & Contractor, D. (2024). On the standardization of behavioral use clauses and their adoption for responsible licensing of ai. arXiv preprint arXiv:2402.05979.

Who is using RAIL?

Total # of models on Hugging Face as of June 12, 2025: 1,785,408

using OpenRAIL licenses:

OpenRAIL license family	37,240
BigScience OpenRAIL-M	1,714
CreativeML OpenRAIL-M	28,613
BigScience BLOOM RAIL 1.0	1,785
BigCode OpenRAIL-M v1	1,431
OpenRail++-M	6,239
Total	77,022

using common open licenses:

Apache 2.0	288,171
MIT	105,358
CC Attribution 4.0	6,618
Total	400,147

How to choose a RAIL?

Use an existing license:

- <u>Samples on RAIL website</u>: could be adapted or modified or used-as-is with minimal changes. Recommended to not modify the usage restrictions without consulting a legal expert to ensure they remain valid legal clauses.

- <u>Licenses listed on Hugging Face</u> (mostly for models)

Use <u>RAIL License Generator</u>: interactive step-by-step tool to create customized licenses. Some commonly used restrictions are included, which can't be removed it seems.

Enforceability

What to do if you or a 3rd party finds a violation of the terms of your license?

1. Reach out to the user

2. Reach out to the platform provider to asses case and if needed take it down

3. Seek legal advice and enforce the license in court, under contract or IP infringement claims.

Critics

Free Software Freedom 0 & OSI's 6th criterion: <u>The freedom to run the program as you wish, for any purpose</u>

<u>Arguments by Richarch Stallman for Freedom 0</u>:

- Restrictions for illegal use is superfluous

- Restrictions for harmful legal use, e.g., torture by security forces will not be respected by states and businesses with powerful lobbies

- Not clear if restrictions would be enforcible

- People have very different ethical ideas, e.g., *"I am not a pacifist, I would also disagree with a "no military use" provision. I condemn wars of aggression but I don't condemn fighting back"*

Challenges & Limitations

- Adoption: RAIL licenses are being adopted at a relatively large scale

- Incompatibility with Open Source: some open source licenses such as GPL V3 forbids adding restrictions that were not originally present

- Enforcement: difficult to enforce by AI developers with limited means. Potential solution: collaborative communities + pro-bono legal initiatives.

- Customization: necessary but can be confusing. Need *standardization*. Good tradeoff: *standardized customization*

Similar licenses

Software licenses with use restrictions:

- Hippocratic License by the Organization for Ethical Source with Corporate Accountability Lab: Allow open-source use *except* for activities that violate human rights, as defined in the UN Universal Declaration of Human Rights.

- Anti-996 License by 966.ICU: Prohibits use by companies that violate labor laws or impose exploitative work schedules (named after China's "996" work culture—9am–9pm, 6 days a week). Derived from MIT license, but adds labor condition restrictions

Data-specific licenses (not use restrictions):

- Community Data License Agreement (CDLA) by The Linux Foundation
- Montreal Data License (MDL) given in a paper by Mila & Element AI researchers + License Generator (domain has expired though)

References

- RAIL Initiative website: https://www.licenses.ai/

- Hugging Face Blog: OpenRAIL: Towards open and responsible AI licensing frameworks

- OECD.AI blog post: Responsible AI licenses: a practical tool for implementing the OECD Principles for Trustworthy AI

- Contractor et al, FAccT 2022 paper: Behavioral Use Licensing for Responsible AI

- McDuff et al, ArXiv 2024 paper: On the Standardization of Behavioral Use Clauses and Their Adoption for Responsible Licensing of AI

- Widder et al, FAccT 2022 paper: Limits and Possibilities for "Ethical AI" in Open Source: A Study of Deepfakes